

Health care expenditures of self-employed farm households in the United States

Ashok K. Mishra*, Hisham S. El-Osta, Mary C. Ahearn

Department of Agricultural Economics and Agribusiness, 211 Woodin Hall, Baton Rouge, LA 70803, USA

Received 6 April 2009; received in revised form 24 September 2010; accepted 4 July 2011

Abstract

For self-employed individuals and their families, purchases of health care services and health insurance policies have the potential to impact their health status, as well as the financial viability of their businesses. Most people in the United States receive health insurance coverage through employer-sponsored programs. Self-employed individuals and their households, such as farm households, may face a greater challenge in getting affordable health insurance. Using a large cross-sectional farm household level dataset, we estimate the impact of the source of health insurance on health care expenditures of farm households in the United States. Results suggest that farm households purchasing individual health insurance directly from vendors are likely to spend more on health care than those with other sources of health insurance. After controlling for a variety of personal and local area characteristics, having health insurance was negatively related to total health care expenditures. Age and income, not surprisingly, were also found to be significant in explaining health care expenditures.

JEL classifications: I10, J22, Q12

Keywords: Health insurance choice; Health care expenditures; Sources of health insurance coverage; Agricultural Resource Management Survey; Two part model

1. Introduction

News sources regularly report on the rise of health care expenditures and consequent limitations on access to healthcare across the country. A recent (2007) national survey found that 41% of working-age adults, or an estimated 72 million people, reported problems paying medical bills and/or the accrual of medical debt, up from 34%, or 58 million adults, only 2 years prior (Doty et al., 2008). Recognizing the importance of health care to constituents, candidates for elected national offices have offered up alternative plans for reforming the health care system to improve affordability and access.

Unlike the majority of developed countries, the underpinning of the U.S. system is employment-sponsored insurance. Under employment-based insurance, health insurance is offered to an employee as part of a compensation package. All of the employees in a firm are considered to be in a group and, therefore, their risks are pooled resulting in lower premiums to all in the pool. However, a number of factors can make individuals ineligible for group insurance programs, including self-employment.

Self-employed individuals often rely on individual policies purchased directly from insurance providers. Usually the premiums are based on the specific characteristics of the individuals covered, such as age and health status. The cost of these premiums for the self-employed population is fully deductible for tax purposes, amounting to an estimated five-year total tax benefit of \$24.3 billion for all self-employed (U.S. Congress, 2008).

In a system relying so heavily on employment-based insurance, U.S. farmers are among the disadvantaged. Moreover, the farming environment has some unique features that place workers and their households at greater risk for injury and illness than many other work environments;¹ however, epidemiological evidence exists that shows farmers have lower rates of many types of deleterious health conditions (Acquavella and Olsen, 1998; Blair et al., 1993). Farm households in the United States can obtain coverage from off-farm jobs of operators or spouses, through their farming operation, government programs, or through the direct purchase of individual plans.

¹ Farming has one of the highest fatality rates of all occupations, according to the U.S. Department of Labor, which reports fatality rates for workers based on their major occupation. While the overall fatality rate in the United States in 2006 was 3.9 per 100,000 workers, the rate for those with farming or ranching as a major occupation was more than nine times that rate—37.1 per 100,000.

*Corresponding author. Tel.: 225 578 0262; fax: 225 578 2716.
E-mail address: amishra@agcenter.lsu.edu (A. K. Mishra).

Considering farmers' lower reliance on employer-sponsored insurance, unraveling the determinants of farmers' expenditures on health care (insurance premiums and out-of-pocket expenses) may yield lessons relevant to the national debate on health care reform.

The specific objectives of this study are twofold: first, we will investigate if there is a difference in the effect of sources of insurance coverage on total health care expenses (insurance premiums and out-of-pocket expenses) of farm households in the United States via the source of the health insurance coverage. Second, we will evaluate the role of some relevant socioeconomic factors such as age, education, size of farm household, and income of farm household on total health care expenditures.

2. U.S. and farm household health expenditures and insurance coverage

In 2007, about the same share—15%—of persons in the U.S. general population and the 5.4 million persons in farm operator households had no form of health insurance (USDA, ERS, 2008b). Although a farm business does not generally offer employment-based health insurance, other factors help to equalize insurance coverage. For example, having health insurance is closely associated with a person's age and income—coverage increases with both, and farm operator households are more than three times as likely to be headed by an individual over 65 compared to other U.S. households. Farm operator households also have higher incomes, on average, than the general U.S. population. Although farm operators are largely self-employed, in the majority of farm operator households, the operator or spouse is also employed off the farm. As with the general population, the most common source of health insurance for members of farm households is employment based. In fact, farm household members are almost as likely as the general U.S. population to receive their health insurance through an outside employer (57.0% and 59.7%, respectively, in 2006). Farm household members are more likely than the general population to directly purchase their health insurance, 20.8% and 9.1%, respectively, in 2006. Further, farm households are less likely (18.5% compared to 27.0%) to receive health insurance from a Government-sponsored program, such as Medicare, Medicaid, or the Veterans Administration (U.S. Bureau of the Census, 2008; USDA, 2008b).

Though farm households have incomes and net worth higher than average U.S. households, much of their net worth is tied up in their farming business and income varies greatly from year to year. For many farm families, health care expenses have the potential to affect not only their families' economic security, but also the financial viability of their businesses. Given their relatively high net worth and reliance on family labor to operate the farm business, providing adequate health insurance to household members is as important to the financial security of the farm business as it is to the health of the family. Opera-

tors of small and medium sized farms are most likely to have family members working off the farm, and many receive fringe benefits such as employer-sponsored healthcare and retirement savings. However, operators and families associated with large farms whose main occupation is farming are less likely to work off the farm and therefore more likely to rely on private health insurance (directly purchased from the vendor).

Historically, the average farm household has had total household expenditures less than the average for all U.S. households, though there is also a difference in the distribution, by income level because of saving and dissaving in the management of the highly variable farm income source (USDA, ERS, 2008b). In contrast, farm households have higher than average expenses for health care. Part of the reason for this lies in the generally older ages of farm operators. To control for age, we examine the expenses and insurance coverage of nonelderly families in Table 1. According to the Medical Expenditure Panel Survey (MEPS), the average U.S. family spent \$1,161 on out-of-pocket health care services and \$1,235 on insurance premiums in 2006 (USDHHS, 2009). This compares to \$2,138 and \$2,969, respectively, for farm households based on the Agricultural Resource Management Survey (ARMS). What accounts for the significantly higher out-of-pocket expenditures of farm households? Part of the reason lies in the source and type of insurance coverage of the populations.

When we consider the nonelderly families, which have all members with private insurance, we see that nonelderly farm households spend more than 40% more out-of-pocket on health care than the general U.S. population. It is also interesting to note that (1) the uninsured farm families spend three times more on out-of-pocket health care services than do the uninsured families in the general U.S. population and (2) in contrast to the general U.S. family population, uninsured farm families spend more out-of-pocket for health care services, excluding insurance premiums, than do farm families where all members are insured. What accounts for this relatively high expenditure in uninsured farm families on out-of-pocket expenses for health care services? If they can afford this level of expenditure, why do they not purchase health insurance or are reasonably priced individual plans not available?² Clearly, there are substantial differences in the health care spending patterns of farm households compared to the general U.S. population, which raises questions about the determinants of their health care spending.

3. Literature review

Research on health insurance, health care expenditures, and the labor market is extensive. We refer the interested reader

² The higher expenses of uninsured farm families may indicate that they are receiving more health care. However, it may also be that the U.S. uninsured family population has greater access to "free" care available in urban centers but not available in rural areas where most farm families reside. An answer to this question is beyond the scope of our farm data.

Table 1

Premium and out-of-pocket health care expenditures among the U.S. nonelderly population and the nonelderly farm operator household population, for selected insurance coverage groups, 2006

| U.S. nonelderly population | | | U.S. nonelderly farm population | | | |
|--|-------------------------|--|---|-------------------------|--|--|
| Insurance status of families (including 1-person families) | Percent of families (%) | Average expenses for health care services (\$) | Average expenses for insurance premium (\$) | Percent of families (%) | Average expenses for health care services (\$) | Average expenses for insurance premiums (\$) |
| All members covered with full-year private insurance* | 54 | 1,410 | 1,981 | 77 | 2,180 | 3,279 |
| Partial private insurance* | 17 | 1,102 | 908 | 9 | 2,219 | 3,493 |
| All members covered with full-year public insurance; no private insurance* | 7 | 643 | NA | 3 | 1,764 | NA |
| Partial year or partial household with public insurance; no private insurance* | 7 | 949 | NA | 1 | 1,321 | NA |
| All members uninsured all year | 15 | 663 | NA | 11 | 2,305 | NA |
| Total | 100 | 1,161 | 1,235 | 100 | 2,138 | 2,969 |

Sources: For U.S. population, USDHH, AHRQ, MEPS data, Tables 1 and 8. For farm population, USDA, 2006 ARMS. NA = not applicable.

Note: Elderly population are included if the head of the household is nonelderly.

*For ARMS, coverage is measured for full-year or part-year coverage combined.

to the excellent surveys by Currie and Madrian (1999) and Gruber and Madrian (2002). The connection of health insurance provision and the workplace setting has important implications not only for the functioning of the U.S. labor market, but also for the rural labor market and farm households, in particular. For example, high health insurance cost has implications for wage determination, employment, and hours' determination in labor market equilibrium (Currie and Madrian, 1999; McDonnell and Fronstin, 1999). Gruber and Poterba (1994) investigated the impact of the 1986 Tax Reform Act, which introduced a new tax subsidy for health insurance purchase by the self-employed, on changing patterns of health insurance demand. The authors found that a 1% increase in the cost of insurance coverage reduces the probability that a self-employed single person will be insured by nearly 2%.

Many studies in the agricultural economics literature have investigated the economic well-being of farm households (Mishra et al., 2002). These studies generally focused on the monetary measure (income and wealth) of well-being. However, the health of individuals and their families is also central to the overall well-being of households. Healthy individuals and families have greater ability to learn new skills, earn more income, and generate wealth to support current and future consumption of individuals and families.

A small amount of literature examines health care issues unique to farm households. The interest in the issue of health insurance among farm families has gained some steam in the early part of 21st century. For example, McNamara (2001) and McNamara and Straub (2002) have provided a descriptive statistical analyses of health insurance coverage and access to hospitals and medical facilities in rural Illinois. More recently, the Access Project, funded in part by the Kellogg Foundation, has conducted a multistate project on health insurance coverage, medical debt, and health care expenditures of farm households.

In their study covering seven Midwestern states, they found that a remarkably high 91% of farm households (with a head less than 65) reported that all members were covered by health insurance for the full year prior to reporting (The Access Project, 2008). However, they also found that the median amount households spent on health care was \$6,700—significantly higher than the general U.S. population. Furthermore, for those households that purchased insurance directly from a private vendor, the median medical expenditure was \$11,200.

Recently, Zheng and Zimmer (2008) investigated farmers' healthcare consumption given their insurances status. They used MEPS data (1996–2001) to identify individuals who indicated that farming was their major occupation (i.e., the “current main job”). The authors pooled the years and treated them as a cross-sectional sample. This method yields a small sample of 261 farmers between the ages 18 and 64. According to their sample, 19% of farmers are uninsured, lower than the national average among self-employed individuals. On the other hand, Zheng and Zimmer point out that 77% of farmers were privately insured, with approximately 69% of privately insured farmers covered through work-related plans. However, due to lack of data on farm work and/or off-farm work (either farmer or spouse) the authors were not able to distinguish the impact of off-farm work on health insurance expenditures. The authors find that farmers with insurance consume significantly more health care (measured as the sum of household expenditures and the amounts reimbursed to providers from insurance companies) than farmers without insurance. Further, their results also show that, on average, uninsured farmers have fewer visits to the doctor or medical provider than an insured farmer per year.

This study has similarities to the recent health care expenditure analysis of U.S. farmers by Zheng and Zimmer (2008), but differs in several fundamental ways. First, we use a farm household model to assess the impact of insurance status on household

health care expenditures. Therefore, we use the information on both farm operator and spouses and other socioeconomic, demographic, and regional factors affecting insurability and health insurance premiums of farm households. The household approach also means that we consider a household insured if all persons in the household reported having coverage. Second, our sample is 10 times larger than Zheng and Zimmer (2008), representing more than 952,442 farm households. Other than the sample of farm households, it is not clear what population is represented in the Zheng and Zimmer study. The MEPS data are not designed to be nationally representative by occupation, and the descriptive statistics of the farmer sample from MEPS differs in significant ways from a representative sample of U.S. farmers. Third, our study utilizes the information on the source of insurance. In particular, we have information on five sources of insurance.

Finally, our fundamental concentration differs from Zheng and Zimmer (2008). We are focusing on the determinants of health care expenditures, including strictly out-of-pocket expenditures, such as household expenditures on health care services and insurance premiums and excluding the payments made by insurance companies to health care providers and the insurance premiums costs paid by employers in employer-sponsored plans. Although economic theory would indicate that workers trade-off direct wage compensation with nonwage benefits, such as employer's share of insurance premiums, empirical work has shown that households are responsive to out-of-pocket expenditures for insurance, rather than the combined employer–employee share of insurance costs. Zheng and Zimmer (2008) measured health care expenditures as the sum of out-of-pocket household expenses paid to health care providers and reimbursements paid by insurance companies to care providers. Their measure of health care expenditures is designed to address their purpose of measuring the role of insurance coverage in the utilization of health care.

4. Data

This study uses the 2006 ARMS. ARMS is conducted annually by the Economic Research Service and the National Agricultural Statistics Service (USDA, ERS, 2008a). The survey collects data to measure the financial condition (farm income, expenses, assets, and debts) and operating characteristics of farm businesses, the cost of producing agricultural commodities, and the well-being of farm operator households. The target population of the survey is operators of farm businesses representing agricultural production in the 48 contiguous states. A farm is defined as an establishment that sold or normally would have sold at least \$1,000 of agricultural products during the year. Farms can be organized as proprietorships, partnerships, family corporations, nonfamily corporations, or cooperatives. Data are collected from one operator per farm, the senior farm operator. A senior farm operator is the operator who makes most of the day-to-day management decisions. For the purpose

of this study, those operator households organized as nonfamily farms were excluded. Nonfamily farms are those where less than half of the farm assets are owned by the principal operator and his or her extended family, whether they reside in the same household or not.

In addition to collecting the standard farm and household information, the 2006 ARMS also queried farmers on two issues related to health-related expenses. First, the questionnaire asked farmers about the total amount of health and/or dental insurance expenses. Second, it asked the respondent about their out-of-pocket medical expenses and health insurance premiums not covered by their insurance plans. The 2006 ARMS data also collected information on the source of health insurance for individuals living in the household. The farm operator was queried on the number of individuals living in the household and how many had health insurance coverage. The choices for health insurance coverage were: (1) off-farm employer of operator; (2) off-farm employer of spouse; (3) farming operation; (4) private purchase; (5) government-provided, such as Medicare or Medicaid.

Since we are interested in health insurance, we restrict our data to households where both the operators and their spouses (some of whom may also be operators) are under the age of 65. All individuals aged 65 or older have the option of receiving Medicare insurance coverage, and virtually all do enroll in the program that is administered by the U.S. government. The data only includes married couples. As a result, the 2006 ARMS yielded 3,292 observations that represent a population of 952,442 farm households with 3.1 million household members. Just over half (55%) of the households had individuals other than an operator and a spouse living in the household; most of the other household members (75%) were children under 18 years old and only 3% were individuals 65 years old or more.

Our data show that about 88% of the individuals associated with the population of farm households we described above had health insurance in 2006. Approximately, 30% of individuals received health insurance through the farm operator's off-farm work, compared to about 18% receiving insurance from the spouse's off-farm work. Only 4% of farm household members in our population were insured through the farming operation. Finally, 15% of farm households purchased health insurance coverage directly from the vendors and 21% of farm households were found with multiple health insurance plans (through Medicare, Medicaid and other forms of public insurance).

The fact that the ARMS data have a complex survey design and is cross-section raises the possibility that the variance of the residual errors is not constant. Accordingly, all standard errors were adjusted for heteroskedasticity using the Huber–White sandwich robust variance estimator (see Huber, 1967; White, 1980). This type of adjustment for standard errors was used in the regression models in lieu of the Jackknife variance estimation method. The Jackknife is a method suitable for estimation of standard errors when the dataset has a complex survey design (for further detail in the context of the ARMS, see Kott 1997;

Dubman 2000), and when the dataset is used in full rather than as a subset.

5. Empirical model

The theory of risk has been used in the health literature to model health insurance demand decisions³ (Arrow, 1963; Feldstein, 1973). In this application, the theory suggests that under conditions of consumer rationality and risk averseness, the decision to purchase insurance is made based on expected utility gain. The utility gains expected from the purchase of health insurance are related to the expected medical need of the person living in the household. For example, Hopkins and Kidd (1996) suggest that the probable distribution of a future health state is based on present and past health states. According to Phelps (1973) uncertainty arises due to the health status of individuals or members of the household. Let us consider a farmer with a utility function $U = U(C, H)$, where C is a “market basket” of consumption goods and H is flow of “healthy days.” Healthy days (H) vary with quantity and quality.⁴ Let us assume that to have healthy days the farmer can buy a contract (or insurance) at price P_h per unit of h . Following a technological production function, $f(h)$ will specify the relationship between the amounts of medical care (h) purchased and the amount of healthy days (H) obtained. Finally, $f(h)$ will have normal production function attributes ($f'(h) > 0$ and $f''(h) \leq 0$). Therefore, the final consumption of health can be written as

$$H = H_0 - \delta + f(h), \quad (1)$$

where δ is a random variable like illness, and $f(h)$ is the production function transforming medical care into healthy days. Equation (1) implies that farmer's consumption of health can be expressed as the sum of initial level of health (H_0) less any random losses, δ , plus any medical care inputs purchased. The farmer's budget is

$$Y = P_c C + P_h h. \quad (2)$$

Inverting Eq. (1) establishes the derived demand for h , given H_0 and δ . Note that demand for h is always conditioned by a specific level of loss, δ . Substituting the value of h into Eq. (2) we can obtain the following constraint for farmer in C and H :

$$Y = P_c C + P_h f^{-1}(H - H_0 + \delta). \quad (3)$$

One can differentiate Eq. (3) (iso-quant) with respect to H to conclude that more C can be acquired only by giving up some H and vice versa. Now consider that the farmer can buy a health insurance policy. The farmer is given the choice of selecting any

coinsurance rate (K) between zero and one; K is the fraction of the medical bills the farmer will pay during the period of the contract. The farmer pays $K P_h$ per unit of h and the insurer will pay $(1 - K) P_h$ per unit. Once the farmer's purchases exceed h^* , the predetermined level of medical services or units of medical care, the insurance policy is no longer effective. Essentially the farmer pays P_h per unit for all $h > h^*$. The insurance premium (I) is in part determined by the particular K chosen by the farmer. The premium may be written as

$$I = I(K, h^*, \theta, \gamma), \quad (4)$$

where θ is a loading charge on the insurance and γ is a vector of parameters that influence the premium (age, family size, education, occupation, etc.) insurance reflects expected expenditures of the farmer, and is represented by

$$I = (1 + \theta) \int_0^{\delta^*} (1 - K) P_h h(\delta) f(\delta) d\delta, \quad (5)$$

where δ^* is exact loss such that $f(h^*) = \delta^*$. Equation (5) holds the appropriate definition of the insurance premium. Now the decision to purchase medical care (h) and other goods (C) follows a normal economic analysis. The farmer is assumed to maximize utility from goods, subject to a constraining income. The farmer can alter his budget line in advance knowing which illness (δ) occurs. Once the probability of $f(\delta)$ is known, or the farmer can allocate the budget in the normal fashion, taking into account the size of the loss⁵ from his health level (H). The budget constraint after the farmer purchases an insurance policy can be represented as

$$\begin{aligned} Y - I &= P_c C + K P_h h & \text{for } h \leq h^*, \\ Y - I &= P_c C + P_h h - (1 - K) P_h h^* & \text{for } h > h^*. \end{aligned} \quad (6)$$

The farmer derives income from two sources.⁶ First, the farmer works F hours on the farm at the W_f wage rate. Second, the farmer works off the farm O hours with wage rate W_o , the total market participation time, T being determined exogenously (T is the sum of farm, off-farm work, leisure time, and time spent to purchase a unit of h). Thus, the total income can be represented as

$$Y = W_f F + W_o O. \quad (7)$$

To choose the optimal C and h given the loss of δ , income, the production function $f(h)$, and market prices P_c and P_h the farmer must maximize the Lagrangian expression with respect to C , h , and λ . Specifically,

$$\text{Max}_{C, h, \lambda} Z = U(C, H) + \lambda(-Y + P_c C + K P_h h + I) \text{ for } h \leq h^*, \quad (8)$$

³ The model presented here does not address the joint decision making that occurs when an employer has the decision to offer insurance and an employee has the decision to take up insurance.

⁴ Flow of services H is subjected to random losses, implying that a rational consumer occasionally becomes ill.

⁵ The insurance policy is chosen before the actual drawing from $f(\delta)$, illness distribution, so that K and h^* can be assumed to be fixed for the consumption decision.

⁶ For simplicity, we only use farmer, although one can expand the model to include the spouse (time and income generated) into these equation as well.

or if the computed value of h is above h^* , then farmer uses

$$\begin{aligned} \text{Max}_{C,h,\lambda} Z = & U(C, H) + \lambda(-Y + P_C C \\ & + P_h h + I - (1 - K)P_h h^*) \quad \text{for } h > h^*. \end{aligned} \quad (9)$$

One can solve the first-order conditions for both Eqs. (8) and (9) represented above. Finally, one can derive the demand function. Specifically, for any loss δ , there is an implied pair of demand functions for C and h of the following general form:

$$C = C(P_C, P_h, W_o, W_f, K, h^*, Y, I, H_0, \delta, \Phi, \Psi), \quad (10)$$

$$h = h(P_C, P_h, W_o, W_f, K, h^*, Y, I, H_0, \delta, \Phi, \Psi), \quad (11)$$

where Φ is a vector of parameters of the utility function and Ψ is a vector of parameters including age, occupation, education, family size, etc., that influence the function $f(h)$. Finally, it should be pointed out that C and h are dependent not only upon income and market prices, but also upon observed loss δ , the entering level of health, H_0 , and the insurance parameter previously chosen. In this study, we are interested in estimating Eq. (11), with the caveat that, due to lack of data, some of the variables, for example, health status, are not available.

6. Estimation method

Health economists often face econometric challenges posed by nature of the health care expenditures data (nonnegative, zero values, and skewness). Alternatives to ordinary least squares (OLS) include a two-part model (2PM) (Blough et al., 1999; Duan et al., 1983, 1984; Hay and Olsen, 1984; Hornbrook and Goodman, 1995; Manning et al., 1981, 1987), which models the probability of nonzero costs separately from their level conditional on nonzero costs.⁷ Mullahy (1998) points out that the choice of modeling health care expenditures, a one-part or two-part model, should be confronted squarely in application. Buntin and Zaslavsky (2004) compare the performances of eight alternative health care cost estimators, including the two-part model, and conclude that a 2PM would be appropriate if modeling nonzero costs. The authors pointed out that the 2PM best predicted the sample mean and had the lowest mean squared forecast error (MSFE) in the split-sample cross-validation exercise. More recently, Meyerhoefer and Pylypchuk (2008) using a 2PM studied the impact of participation in food stamp programs on female obesity and health care spending. The authors point out that 2PM has been shown to outperform more general sample selection models in the common case where an exclusion restriction to empirically identify the first stage is not available.

In this article, the impact of farm households' health insurance status on health care expenditures (total premiums and

out-of-pocket expenses) is estimated using the two-part model. In the first part of the model, it is assumed that the probability (P_i) of the i th farm household having a nonzero health care expenditure is governed by a parametric binary model as follows:

$$\begin{aligned} Y_i = 1 & \quad \text{if } Y_i^* = \beta' X + v_i > 0, \\ = 0 & \quad \text{if } Y_i^* \leq 0 \text{ (otherwise),} \end{aligned} \quad (12)$$

where Y_i^* is an unobservable random variable since it is derived from a farm operator's own utility function. The expected value of Y_i is estimated using the following probit regression model:

$$\begin{aligned} E[Y|X] &= P(Y_i = 1) \\ &= P(Y_i^* > 0) = P(-v_i < \beta' X_i), \end{aligned} \quad (13)$$

Using a maximum likelihood procedure to estimate (13) allows for the computation of the following (see Greene, 2008):

$$\begin{aligned} \hat{P}_i &= \Phi(\hat{z}_i) = \int_{-\infty}^{\hat{z}_i} \varphi(u_i) du_i \\ &= \int_{-\infty}^{\hat{z}_i} (2\pi)^{-1/2} \exp(-u_i^2/2) du_i. \end{aligned} \quad (14)$$

where $\Phi(\cdot)$ is the standard cumulative distribution function, $\varphi(\cdot)$ is the probability density function of the standard normal, u_i (equivalent to $-v_i$ in (13)), which is redefined to keep the algebra simple) is a random variable with mean zero and unit variance, and $\hat{z} = \hat{\beta}' X$.⁸

In the second part of the 2PM, the following OLS model of total out-of-pocket health care expenditures (y) is estimated:

$$E[\ln(y)|y > 0; x] = \alpha' x + E(\varepsilon_{OLS}|y > 0; x) = \alpha' x. \quad (15)$$

The logarithmic transformation in (15) is undertaken to mitigate the ill-effects on estimated regression parameters when the distribution of the dependent variable, as in this article, is skewed and with a long right tail. Many studies can be found in the literature that have discussed the merits of transforming the data using the logarithm function as a possible fix to the skewness problem (e.g., Manning, 1998; Manning and Mullahy, 2001; Mullahy, 1998).

Calculation of the marginal effects of the explanatory variables in (15) under the assumption of normal and homoskedastic errors ε starts by considering the possibility of $P_i(y = 0)$, which represents the likelihood that the farm household has no health care expenditures. In turn, prediction of y consists of two parts: $P_i(y > 0)$ which is estimated from the first part of the 2PM using the probit regression model, and the conditional expectation

⁷ Data structures are observed in health applications such as health care utilization, expenditures, use of tobacco and alcohol, physician service, etc.

⁸ The marginal effects (Greene, 2008) are computed as in $\frac{\partial E[Y|X]}{\partial X} = \varphi(\hat{\beta}' X)\hat{\beta}$.

$E[y|y > 0]$ from the log-linear OLS regression model (see Dow and Norton, 2003; Frondel and Vance, 2010):⁹

$$\begin{aligned} E[y] &= P(y > 0)E[y|y > 0] + P(y = 0)E[y|y = 0] \\ &= P(y > 0)E[y|y > 0] + 0 \\ &= \Phi(\hat{\beta}'X) \exp\{\hat{\alpha}'x + 0.5\sigma^2\}. \end{aligned} \quad (16)$$

Utilizing the fact that the derivative of the cumulative normal function Φ equals the normal density function ϕ , and using the product and chain rules of differentiation allows for the derivation of the marginal effect (ME), and consecutively after dividing ME by $E(y)$ and multiplying it by x_k for the derivation of the elasticity (η) of the k th continuous explanatory variable as in (see Frondel and Vance, 2010; Norton et al., 2008)

$$\begin{aligned} ME_{x_k} &= \frac{\partial E[y]}{\partial x_k} = \hat{\alpha}_k E[y] + \hat{\beta}_k \Phi(\hat{\beta}'X) \exp\{\hat{\alpha}'x + 0.5\sigma^2\} \\ &= \hat{\alpha}_k E[y] + \beta \frac{\phi(\hat{\beta}'X)}{\Phi(\hat{\beta}'X)} E[y], \end{aligned} \quad (17)$$

$$\eta_k = \frac{\partial \ln E[y]}{\partial \ln x_k} = \frac{\partial E[y]}{\partial x_k} \frac{x_k}{E[y]} = \left[\hat{\alpha}_k + \beta_k \frac{\phi(\hat{\beta}'X)}{\Phi(\hat{\beta}'X)} \right] x_k. \quad (18)$$

When the k th explanatory variable in (15) is a dummy variable (D_k), its elasticity is computed, as described in Frondel and Vance (2010), as a “relative difference” in the expected value of y as in

$$\eta_{D_k} = (E[y|D_k = 1] - E[y|D_k = 0])/E[y|D_k = 0]. \quad (19)$$

An important consideration for the robustness of retransformations in the estimation of 2PM in health economics is the issue of whether the distribution of $y|y > 0$ is log-normally distributed with a constant σ^2 . If this were the case indeed, then use of $E(y|y > 0) = \exp(\hat{\alpha}'x + 0.5\sigma^2)$ as suggested in (16) is hence justifiable. As a robust alternative when this assumption is violated, as was suggested by Duan (1983), is to replace $E(y|y > 0) = \exp(\hat{\alpha}'x + 0.5\sigma^2)$ with $E(y|y > 0) = S \exp(\hat{\alpha}'x)$ where the smearing estimator $S = N_+^{-1} \sum_{i \in n_+} \exp(\hat{\varepsilon}_i)$ and where N_+ is the size of the sample n_+ containing positive values of y .

As an alternative to the logarithmic transformation used in the estimation of out-of-pocket expenditures for farm households, and in concentrating on part 2 of the 2PM, we also consider the possible use of a Gamma-based generalized linear model (GLM) with a log-link function as has been the practice in health expenditure applications (see Blough et al., 1999; Buntin and Zalavsky, 2004; Gilleskie and Mroz, 2004)¹⁰

$$\mu = E[y] = E[y|y > 0] = \exp(\hat{\alpha}'x). \quad (20)$$

⁹ Let a random variable $q = \ln(y)$. If q has a normal distribution with some mean value of $E(q) = \mu$ and variance σ^2 , then y will have a lognormal distribution and a mean value of $E(y) = \exp(\hat{\alpha}'x + 0.5\sigma^2)$ (see Shen et al., 2006).

¹⁰ A note with regard to the nature of the sample used in the GLM model is in order. While the GLM model allows the use of all observations in the analysis (i.e., both zero- and nonzero values of y), we restricted our analysis for the GLM model to only those observations where $y > 0$, thus allowing for $E[y] = E[y|y > 0]$ by construct. This was done to allow for comparability of the regression results between the log-linear part of the 2PM and the GLM models. Also, note that the variance structure for the GLM model with link function g is described as $v = g(\mu)^2$.

The marginal effect (ME_{glm}) of the k th continuous explanatory variable and its corresponding elasticity η_{glm} based on this model are computed as follows (see Decker, 2009; Manning et al., 2005; Wooldridge, 2002):

$$ME_{\text{glm}, x_k} = \frac{\partial E[y|y > 0]}{\partial x_k} = \hat{\alpha}_k \exp(\hat{\alpha}'x), \quad (21)$$

$$\eta_{\text{glm}, x_k} = \frac{\partial E[y|y > 0]}{\partial x_k} \frac{x_k}{E[y|y > 0]} = \hat{\alpha}_k x_k. \quad (22)$$

When the explanatory variable is a dummy variable, its elasticity is computed as in (19) with $E[y]$ based on the formulation in (20).

7. Results

Table 2 presents definitions and summary statistics for the variables used in the analysis.¹¹ Note that the mean health expenditures for farm households, including those with zero expenditures in 2006 was \$5,516. For only those farm households who had any health care spending, which accounted for 93% of the weighted sample, the mean health expenditures was a slightly higher at \$5,931. The combined income and wealth measure used in the second part of the 2PM, referred to henceforth as full-income. Here we use 2005 full income in order to mitigate the potential for “inconsistent” parameter estimates due to endogeneity concerns.¹² Fig. 1 presents a map of the rural-urban location codes used in the models.

The upper part of Fig. 2 depicts, based on a spike plot of the weighted data (see Cox, 2004), a skewed distribution of health expenditures with a long right tail. The first leftward spike in the figure shows nearly 67,000 farm households, of the nearly one million farm households in the selected sample, with zero health expenditures. The lower portion of the figure shows the distribution of health expenditures after a log transformation. The distribution becomes better behaved—although it remains

¹¹ A Ramsey regression specification error test (or commonly known as the RESET test; see Ramsey, 1969) for omitted variables (e.g., age², full-income²) was performed. Specifically, applying RESET to examine the null hypothesis H_0 : “Model has no omitted variables” yielded an $F(3, 3007) = 1.25$ ($Prob > F = 0.29$), which thus indicates that the model described in Table 2 is properly specified.

¹² The combined income-wealth (CWB) measure used is similar to what has been proposed initially by Weisbrod and Hansen (1968) and reaffirmed later by Hill (2002):

$$CWB_i = HHMI_i + (tc \text{ MNW}_i) \frac{r}{[1 - (1 + r)^{-l}]},$$

where tc represents a proportional adjustment factor reflecting transaction costs, l is the life expectancy of the unit, r is an assumed interest rate set at 4% for this paper, and $HHMI_i$ and MNW_i represent the i th total household’s money income in 2005 and marketable net worth, respectively. Because the 2006 ARMS did not query farm operators about their levels of farm and nonfarm assets and debts in 2005, the 2006 levels of MNW_i were used instead (with a downward adjustment due to inflation in 2005) under the assumption that a farm household’s total equity in 2005 was the same as in 2006.

Table 2
Definitions and weighted means of variables used in parts 1 and 2 of model of health care expenditures for farm households, 2006

| Definitions | First part ($y \geq 0$) | Second part ($y > 0$) |
|---|---------------------------|-------------------------|
| Dependent variable (y) | | |
| Total health care expenditures on health care (=1; 0 otherwise) | 0.93 | – |
| Total health care expenditures on health care (\$1,000) | 5.516 (0.26) | 5.931 (0.28) |
| Log(total health care expenditures on health care) | – | 1.61 (0.03) |
| Explanatory variables | – | – |
| <i>Operator and household characteristics</i> | | |
| Age of farm operator (years) | 50.65 (0.37) | 50.79 (0.38) |
| High school education (=1; 0 otherwise) | 0.37 | 0.37 |
| Some college (=1; 0 otherwise) | 0.25 | 0.26 |
| College (=1; 0 otherwise) | 0.30 | 0.30 |
| Gender of principal farm operator (=1 if operator is female; 0 otherwise) | 0.07 | 0.08 |
| Ethnicity of principal farm operator (=1 if operator is white; 0 otherwise) | – | 0.97 |
| Size of farm household | 3.22 (0.06) | 3.23 (0.07) |
| Miles from the closest city with a population of 10,000 or more | 24.56 (0.97) | 24.42 (1.06) |
| Combined income and wealth measure, 2005 (\$1,000) | – | 105.831 (4.40) |
| Sources of insurance* | – | – |
| Off-farm employer of principal operator (=1; 0 otherwise) | – | 0.30 |
| Off-farm employer of spouse (=1; 0 otherwise) | – | 0.18 |
| Farming operation (=1; 0 otherwise) | – | 0.04 |
| Private, fully purchased by the household (=1; 0 otherwise) | – | 0.15 |
| Multiple, including Medicare, Medicaid, or other public insurance (=1; 0 otherwise) | – | 0.21 |
| <i>Farm characteristics</i> | | |
| Sole proprietorship (=1; 0 otherwise) [†] | 0.94 | – |
| Full ownership of farming acreage (=1; 0 otherwise) | 0.57 | – |
| Cash grains (=1; 0 otherwise) | 0.11 | – |
| Other crops (=1; 0 otherwise) | 0.24 | – |
| Fruits and vegetables | 0.04 | – |
| Dairy | 0.04 | – |
| Beef cattle and hogs | 0.35 | – |
| Location is in the following county [‡] | | |
| Large metro: population of 250,000 or more (=1; 0 otherwise) | 0.23 | 0.23 |
| Small metro: population of fewer than 250,000 (=1; 0 otherwise) | 0.06 | 0.07 |
| Large urbanized: urban population of 20,000 or more (=1; 0 otherwise) | 0.10 | 0.10 |
| Small urbanized: urban population of less than 20,000 (=1; 0 otherwise) | 0.45 | 0.45 |
| Sample size | 3,292 | 3,028 |
| Farm operator households | 952,442 | 885,868 |

Data source: 2006 Agricultural Resource Management Survey (Version 1, Phase III). Standard errors (SE) are in parentheses. Estimates of SE were measured using the bootstrapping variance estimation method with 1,000 drawn samples. The sample includes the households of married farm operators who are less than 65 years old.

*The excluded category includes farm households with no health insurance.

[†]The excluded category includes all other forms of farm legal organizations such as legal partnerships and corporations. For type of farm the excluded category is livestock and poultry farms.

[‡]The excluded category includes completely rural counties or counties with less than 2,500 urban population.

not quite symmetric due to the presence of a large number of farm households with zero health expenditures—than the distribution based on the raw scale, with the coefficients of skewness and kurtosis dropping from 16.93 to 0.05 and from 50.93 to 2.41, respectively. Fig. 3 demonstrates the inappropriateness of using OLS in the second part of the 2PM in the analysis of the demand for health care expenditures. Specifically, the upper portion of the chart shows that, when OLS is used, there is the presence of heteroscedasticity and negative prediction of health care expenditures.¹³ A scatter plot of log-scale residu-

¹³ The fit line in the upper chart based on a locally weighted regression (lowess) function (for more detail, see Bowman and Azzalini, 1997) alludes to a variance of the raw-scale residuals as a function of the predictions as being

als over the log-scale predictors, as demonstrated in the lower portion of Fig. 3, highlights the benefit of using OLS, with a log-transformed dependent variable. This transformation significantly lowers the coefficients' skewness and kurtosis in the distribution of the raw-scale residuals, from 17.69 to 0.34 and from 544.14 to 3.32, thus making the distribution of the residuals closer to fulfilling the normality assumption as required in OLS estimation.¹⁴

quadratic, thus pointing to the possibility of improvement in the estimation of health care expenditures when a gamma or a logarithmic transformation of y is used rather than when y is modeled using OLS regression without such a transformation.

¹⁴ A plot of the quantiles of the residuals from the log-transformed model against the quantiles of a normal distribution (not included in the paper to

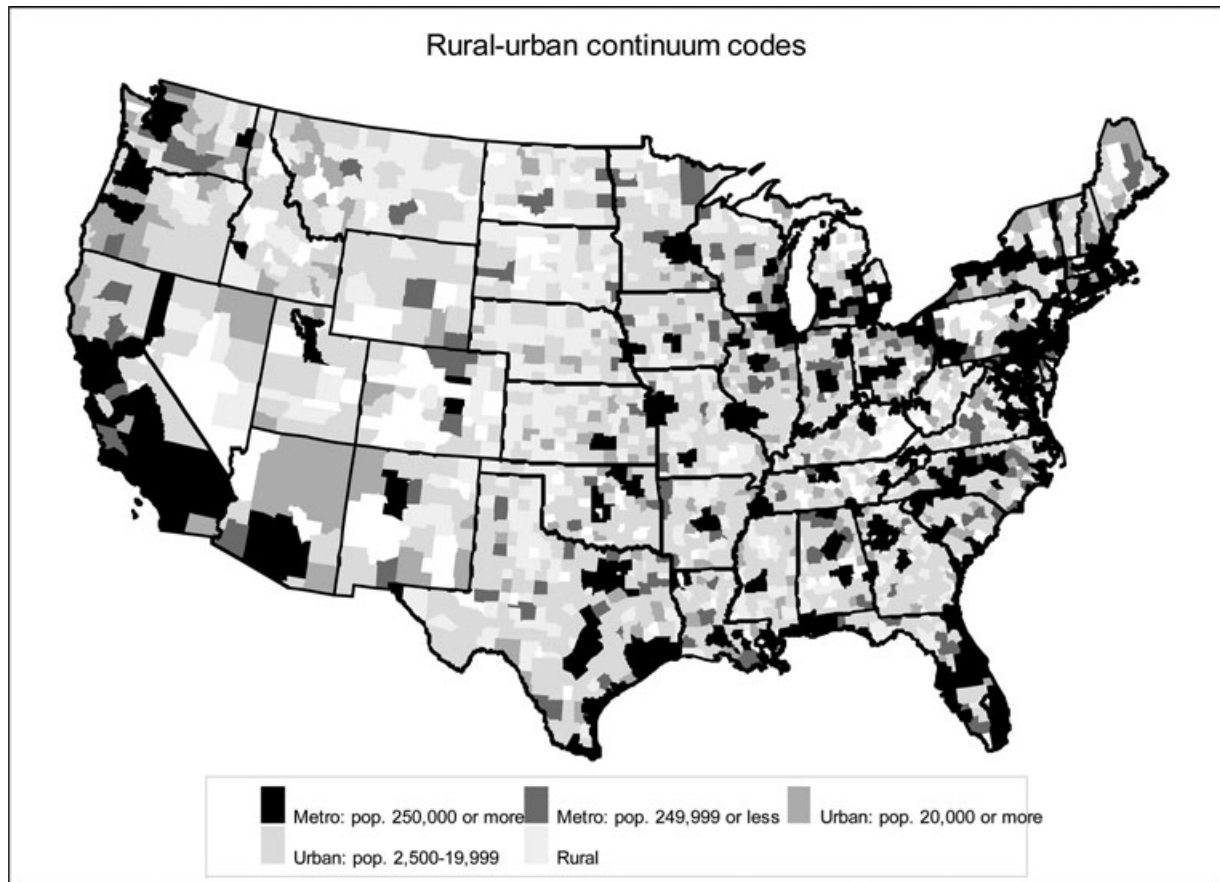


Fig. 1. Rural–urban location codes used in the regression models.

The first column of statistics in Table 3 reports the parameter estimates of the first part of the 2PM, estimated using a probit model. These estimates describe the direction and the magnitude of the effects of the variables on the likelihood that a farm household has expended any money on health care in 2006. Results indicate that the coefficients of the variables depicting the demographic characteristics (age, education, race, and gender) of the farm operator are all significant at the 5% level of significance, or better. Of the remaining variables, only the size of farm households is found to have a significant impact on the likelihood of health care spending.¹⁵ In terms of the marginal impacts of the explanatory variables on the probability

save on space) showed a departure from normality only near the tails. Also, a Breusch–Pagan (BP)/Cook–Weisberg test for H_0 : “Constant variance” using the raw scale dependent variable y yielded a $\chi^2(17) = 3385.87$ ($Prob > \chi^2 = 0.00$), and when using the log-transformed dependent variable y , this yielded a $\chi^2(17) = 62.45$ ($Prob > \chi^2 = 0.00$), thus affirming the presence of heteroskedasticity under both cases (for more detail, see Breusch and Pagan, 1979; Cook and Weisberg, 1983). To the extent that the BP test assumes that the distribution of the error terms is normal, a discernment of the presence of heteroskedasticity is thus more reliable only when y is log-transformed since only then the normality assumption becomes closer to being met.

¹⁵ An expanded probit regression model with sources of insurance as an added explanatory variables was estimated but was not included as part of the analysis due to the endogeneity of these variables. While detailed reporting of the results can be obtained from the authors upon request,

the farm operator having received some college education is found to exert the strongest impact as indicated by the rise in such a probability by nearly 8%.

Results in the first column in Table 4 are those from a log-transformed y using the OLS regression model. They show a positive and significant effect of the age of the farm operator on health care expenditures. Consistent with the economic theory, results indicate that health care expenditures rise with age. Age may act as an important determinant of propensity to insure, not only because it is a variable associated with high indirect risk vulnerability and thus increased expected medical consumption, but also because it is associated with an increased stock of wealth. Stock of wealth generally increases as individuals/households get older. Van De Ven and Van Praag (1981) and Hopkins and Kidd (1996) note that younger individuals and families are generally less well off. Results are consistent with the findings in the literature (Cameron et al., 1988; Ngui et al., 1990; Savage and Wright, 1999).

quest, the following is a shortened description of the probit model’s results (significant [at 5% level or better] coefficients are in bold numbers): $-0.6.05 + \dots -0.137^*$ Off-farm employer of principal operator -0.017^* Off-farm employer of spouse -0.227^* Farming operation $+0.574^*$ Private, fully purchased by the household -0.070^* Multiple other sources $+\dots$ (Pseudo R -squared = 0.086).

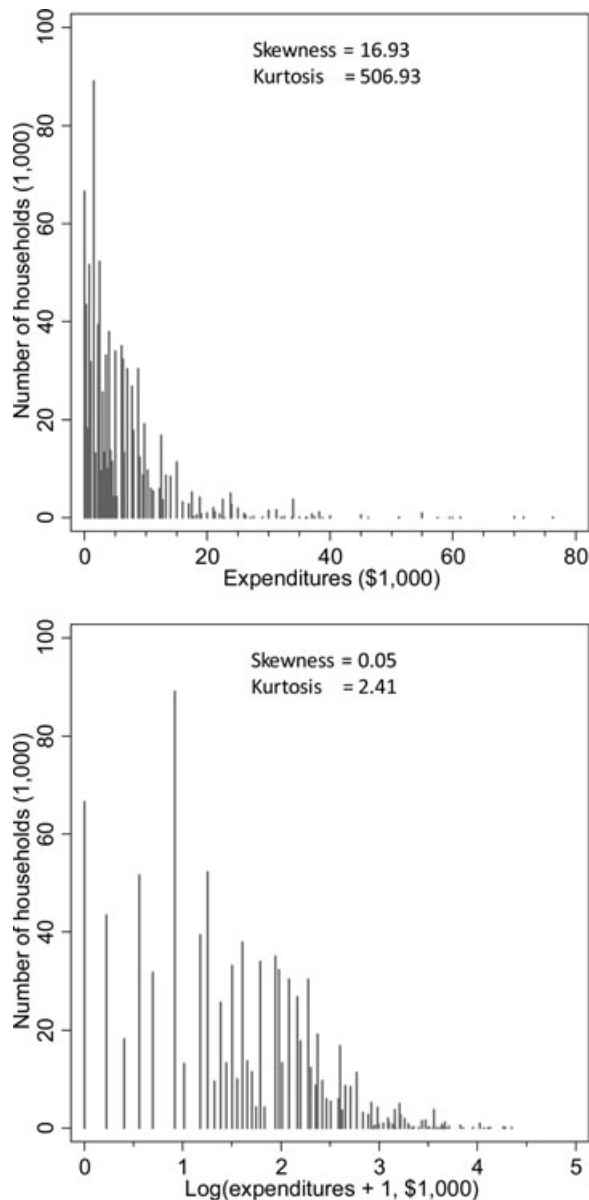


Fig. 2. Distribution of health care expenditures, 2006.

Having some or completed college education (including the possibility of graduate education) is shown to have a positive and significant influence on health care expenditures. Of the variables depicting household characteristics, both “size” and “full-income” of the farm household are significant variables in explaining health care expenditures. The reason for the positive impact of income could be that higher income increases the likelihood of buying health insurance (Propper, 1989; Savage and Wright, 1999), which could result in higher health care expenditures. Higher income generally decreases the opportunity cost associated with the purchase of health insurance in pure monetary terms (Hopkins and Kidd, 1996). Our results are consistent with the findings of Van De Ven and Van Praag (1981) for the United States, as well as with research on health insurance

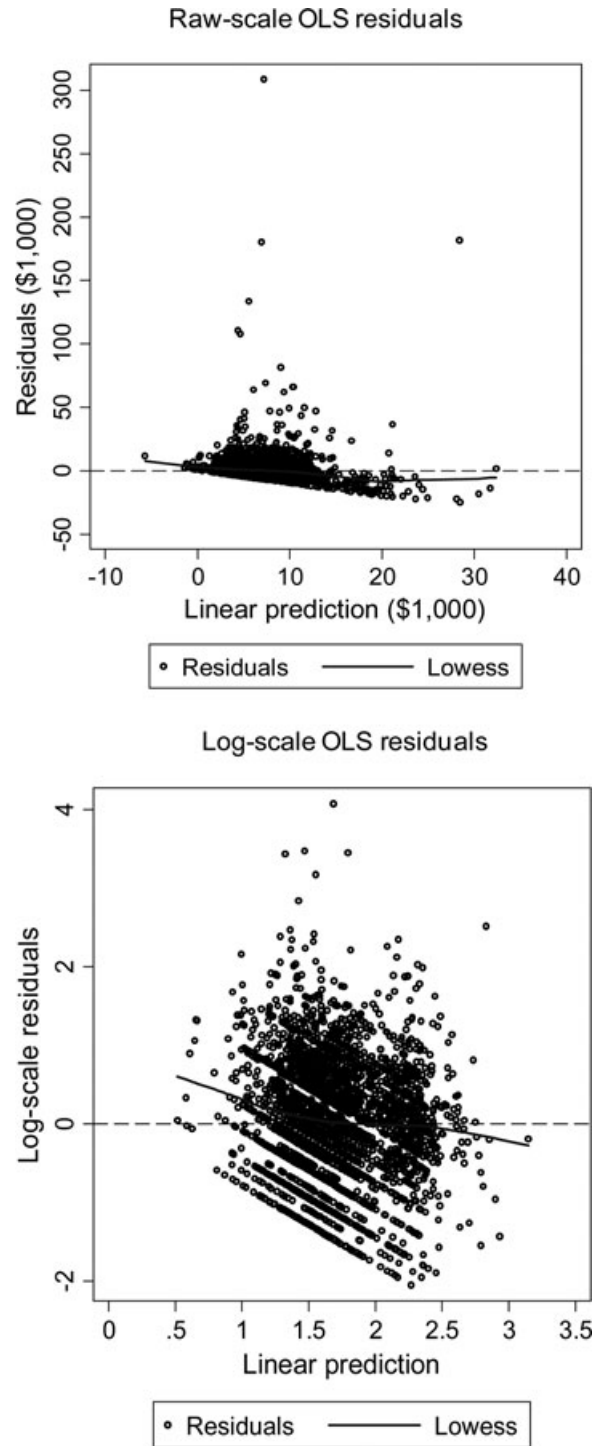


Fig. 3. Heteroscedasticity in raw- and log-scale OLS residuals, 2006.

purchase decisions in studies conducted around the world (for example, Propper, 1989 in the United Kingdom; Cameron et al., 1988 in Australia; and Hurd and McGarry, 1997 in the United States). All of the variables depicting the sources of farm insurance are found statistically significant in terms of their positive impact on health care expenditures, particularly when the farm

Table 3
Probit estimation results of the health care expenditures' decision, 2006

| Variables | $\hat{\beta}$ | Robust standard errors | $\frac{\partial P}{\partial X}$ | Robust standard errors |
|---|---------------|------------------------|---------------------------------|------------------------|
| Intercept | −0.6004 | 0.690 | — | — |
| Age | 0.0203*** | 0.008 | 0.0023** | 0.000 |
| High school education | 0.6346** | 0.290 | 0.0669** | 0.029 |
| Some college | 0.8940*** | 0.286 | 0.0768*** | 0.020 |
| College | 0.6685** | 0.287 | 0.0648** | 0.024 |
| Female | 0.6857*** | 0.211 | 0.0509*** | 0.011 |
| Size of farm household | 0.1229** | 0.054 | 0.0143*** | 0.006 |
| Miles | −0.0018 | 0.002 | −0.0002 | 0.000 |
| Sole proprietorship | −0.0125 | 0.183 | −0.0014 | 0.021 |
| Full ownership | 0.0323 | 0.148 | 0.0037 | 0.017 |
| Cash grains | −0.1395 | 0.215 | −0.0177 | 0.029 |
| Other crops | 0.0824 | 0.207 | 0.0093 | 0.022 |
| Fruits and vegetables | 0.1657 | 0.284 | 0.0171 | 0.026 |
| Dairy | 0.2562 | 0.243 | 0.0248 | 0.019 |
| Beef cattle and hogs | 0.1956 | 0.193 | 0.0219 | 0.020 |
| Large metro | −0.0085 | 0.236 | −0.0009 | 0.027 |
| Small metro | 0.1537 | 0.307 | 0.0161 | 0.028 |
| Large urbanized: urban population of 20,000 or more | 0.1910 | 0.263 | 0.0197 | 0.024 |
| Small urbanized: urban population of less than 20,000 | −0.0971 | 0.198 | −0.0114 | 0.023 |
| Pseudo R^2 | 0.07 | | | |

Data source: 2006 Agricultural Resource Management Survey (Version 1, Phase III).

*Significant at 10%.

**Significant at 5%.

***Significant at 1%.

household directly purchases insurance.¹⁶ Findings indicate no association between the location of the farm where the operator resides and health care expenditures.

Results for the Gamma-based GLM with a log-link function of health care expenditures are reported in column 3 of Table 4. The central structure of this model, as indicated earlier, is an exponential conditional mean (see Eq. (20)) with variance of y that is proportional to the square of its mean (see footnote 9).¹⁷

¹⁶ A caveat is in order here. As one reviewer has correctly pointed out, the dummies depicting the sources of insurance in the health care demand model can be viewed as being endogenous. However, no proper instruments can be found in the ARMS to allow for the use of an instrumental variable (IV) approach in order to mitigate the ill-effects of endogeneity on estimated parameters; these dummies are nevertheless used in the article under the assumption that they are exogenous. It is important to note, however, that issues of endogeneity concerns related to dummies depicting insurance sources in a health care utilization equation were addressed by Cameron et al. (1988), with findings indicating no substantial differences between models that attempted to correct for the endogeneity of these variables using an IV approach and those that did not.

¹⁷ To discern whether the assumption of the exponential conditional mean being proportional to the square of its mean is valid, a modified version of the Park test for heteroscedasticity is used (see Manning and Mullahy, 2001; Park, 1966). Specifically, a regression of the form

$$\ln(y_i - \hat{y}_i)^2 = \lambda_0 + \lambda_1 \ln(\hat{y}_i) + v_i$$

Table 4
Results from the two-part model (2PM) of health care expenditures using OLS and GLM regression procedures, 2006

| Variables | OLS (dep. var.: log(y)) | | GLM (dep. var.: y) | |
|---|-------------------------|------------------------|--------------------|------------------------|
| | $\hat{\alpha}$ | Robust standard errors | $\hat{\alpha}$ | Robust standard errors |
| Intercept | −0.3663 | 0.324 | −0.7274 | 0.461 |
| Age | 0.0165*** | 0.003 | 0.0260*** | 0.004 |
| High school education | 0.2398 | 0.160 | 0.2114 | 0.261 |
| Some college | 0.2756* | 0.160 | 0.2016 | 0.259 |
| College | 0.3217** | 0.157 | 0.2322 | 0.258 |
| Female | −0.1453 | 0.112 | −0.1374 | 0.173 |
| White | 0.3541 | 0.256 | 0.3436 | 0.337 |
| Size of farm household | 0.0796*** | 0.019 | 0.0895*** | 0.028 |
| Previous year's combined income and wealth measure | 0.0003** | 0.000 | 0.0005** | 0.000 |
| Off-farm employer of principal operator | 0.1868* | 0.098 | 0.2553* | 0.141 |
| Off-farm employer of spouse | 0.1983** | 0.099 | 0.2649* | 0.136 |
| Farming operation | 0.4412*** | 0.117 | 0.6209*** | 0.135 |
| Private, fully purchased by the household | 0.8920*** | 0.090 | 0.9505*** | 0.125 |
| Multiple other sources | 0.3903*** | 0.107 | 0.4580*** | 0.142 |
| Large metro | −0.1060 | 0.099 | −0.1554 | 0.141 |
| Small metro | 0.0009 | 0.100 | −0.1426 | 0.132 |
| Large urbanized: urban population of 20,000 or more | −0.0976 | 0.096 | −0.2269 | 0.131 |
| Small urbanized: urban population of less than 20,000 | −0.0688 | 0.081 | −0.1514 | 0.121 |
| R^2 | 0.181 | | 0.052# | |
| F test [degrees of freedom] | 15.09*** | [17, 3010] | | |
| Wald test (χ^2 stat) [degrees of freedom] | | | (165.97)*** | [17] |

Data source: 2006 Agricultural Resource Management Survey (Version 1, Phase III).

#This is the squared correlation coefficient (r^2) where r is the coefficient measuring the extent of the correlation between the observed y and the predicted y .

*Significant at 10%.

**Significant at 5%.

***Significant at 1%.

Findings show a similarity between the results based on the OLS log-transformed model and those that are based on the GLM, except for the college education variables that were statistically insignificant in the latter model. A notable difference among the two models, however, which explains the insignificance of the college-education variables in the GLM model, is the fact that standard errors under GLM tend to be larger. The large loss

is estimated using a robust variance–covariance matrix of estimated coefficients. This test has yielding a statistically significant value of 1.44 for λ_1 . An F -test of this value being equal to 1 was rejected based on an $F(1, 3026) = 5.92$ ($Prob > F = 0.0150$), thus affirming the notion that the variance of the error term can be approximated as being proportional to the square of its mean.

Table 5
Elasticities of health care expenditures for a selected set of covariates, 2006

| Variables | % change in expenditures | | | |
|--|--------------------------|-------------------|--------------------|-------------------|
| | OLS (dep. var.: log(y)) | | GLM (dep. var.: y) | |
| | | Robust Std. Error | | Robust Std. Error |
| 1% increase in continuous variables: | | | | |
| Age | 0.966*** | 0.0030 | 1.309*** | 0.0042 |
| Size of farm household | 0.312*** | 0.0023 | 0.292*** | 0.0022 |
| Previous year's combined income and wealth measure | 0.051*** | 0.0015 | 0.080*** | 0.0023 |
| Relative differences in dummy variables: | | | | |
| Some college | 0.295*** | 0.0007 | | – |
| College | 0.347*** | 0.0009 | | – |
| Sources of insurance: | | | | |
| Off-farm employer of principal operator | 0.200*** | 0.0002 | 0.149*** | 0.0008 |
| Off-farm employer of spouse | 0.213*** | 0.0003 | 0.154*** | 0.0009 |
| Farming operation | 0.523*** | 0.0013 | 0.362*** | 0.0018 |
| Private, fully purchased by the household | 1.255*** | 0.0064 | 0.554*** | 0.0028 |
| Multiple other sources | 0.452*** | 0.0010 | 0.267*** | 0.0013 |

Data source: 2006 Agricultural Resource Management Survey (Version 1, Phase III). Estimates of SE were measured using the bootstrapping variance estimation method with 1,000 drawn samples.

*Significant at 10%.

**Significant at 5%.

***Significant at 1%.

of efficiency witnessed under the GLM model is not surprising (see Buntin and Zaslavsky, 2004; Manning and Mullahy, 2001), particularly when the kurtosis of the distribution of the error terms under the OLS log-normal model is in excess of 3, as in this study where the value of this coefficient is 3.32.

This article has used two different techniques to assess the impact of farm and farm household characteristics, including sources of health insurance, on health care expenditures. Of the two models used, the OLS model with a logarithmic transformation of health care expenditures seems to perform much better than that of the GLM based on goodness of fit measures ($R^2 = 0.181$ versus an $R^2 = 0.052$, respectively), and based on efficiency gains in the measurement of estimated variances of regression parameters. Despite this, and for the sake of providing some sort of sensitivity analysis to the extent of variation in the elasticity estimates between various methods of estimation, the next section computes the elasticities of health care expenditures based on the regression results from both models.

The elasticities of health care expenditures for selected covariates (those that were found statistically significant in Table 4) and under both the OLS model with a log-transformed y and the GLM with a raw-scale y are presented in Table 5

(see Eqs. (18) and (21), respectively). Findings indicate significant statistical differences, due to nonoverlapping confidence intervals, in the estimated elasticities based on the regression method chosen to estimate $E(y|x)$ and consequently the elasticities themselves. For example, the elasticity of health care expenditures with respect to full-income is 0.051 with the log transformation of y and 0.081 without such a transformation. Put differently, a 1% increase in full-income will increase spending in health care by about 5% based on the first model and by about 8% based on the second model. Another equivalent interpretation for these estimates (i.e., $\partial y/y = (100 \hat{\alpha}) \partial x$) put a one unit increase or change in full-income (i.e., \$1,000), which is a covariate that does not enter in the probit model in the first part of the 2PM, to result in 0.03% (or \$226) increase in expected health expenditures based on the model with transformed y and in 0.05% (or \$296) based on the GLM model.¹⁸

The results in Table 5 also show that the elasticities of health care expenditures in terms of sources of health insurance are highest when insurance is purchased directly by the farm household. Under the OLS-transformed model, this elasticity is at 1.255, and is at 0.554 when the estimation of the elasticity is based on the GLM. When translated into dollar amounts, these results indicate that farm households with private health insurance are likely to spend either \$9,479 ($= 1.255 \times \$7,553$; based on the model with transformed y) or \$3,286 ($= 0.554 \times \$5,931$; based on the GLM model) more on health care expenditures than farm households without any health insurance.¹⁹ Table 4 further reports that for farm households who have health insurance, the age of the farm operator and household income are important determinants of health care expenditures. In both cases, health care expenditures are positively correlated with age and income of the farm households. Access to health care facilities and location of households may have an impact on health care expenditures. For example, in many rural areas, a lack of competition among health care providers may result in higher health care expenses. However, our results (in Table 4) show that the location of farm households is not a significant factor in explaining health care expenditures. Part of the explanation for this finding is likely related to farm households in more isolated areas traveling to metro areas to seek care; the cost associated with this additional travel is not included in health care expenditures of the household.

¹⁸ While the expected health care expenditures based on the GLM was based on Eq. (21), the OLS log-transformed model was computed based on Eq. (16) as in the following with some adjustment as to reflect the utilization of the smearing estimator S :

$E(y|y > 0) = S \exp(\alpha'x)$.

¹⁹ However, as one reviewer noted, caution should be exercised when reporting these estimates due to issues related to endogeneity of insurance. It is often difficult to find credible instruments to correct for endogeneity in empirical models. Further, due to lack of data on health status the issue of endogeneity may be amplified. In order to fully address this issue one may need a natural experiment or quasi-experimental design to identify a credible instrument.

8. Summary and conclusions

Unlike the general U.S. population, descriptive data for uninsured farm households indicate that uninsured households spend more for health care services than the insured. When out-of-pocket expenses and health insurance premiums are considered, insured farm households spend more than uninsured farm households. In our model, we found that when we controlled for a variety of personal and local area characteristics, having health insurance was negatively related to total health care expenditures. Age and income, not surprisingly, were also found to be significant in explaining health care expenditures.

Farm groups often express concern over the high price and low availability of individual insurance policies. Given the current debate on health care reform, we also considered the role of the source of health insurance in explaining total health care expenditures. The majority of farm households engage in multiple-job holding and have access to employer-sponsored plans, but still a large share of households have directly purchased private insurance plans. Enrollment in a directly purchased private insurance plan with an insurance vendor was found to be a significant factor in explaining health expenditures. Directly purchased private insurance plans are generally more expensive in terms of premiums, deductibles, and co-pays than other types of plans. With the new health care legislation, a major change in directly purchased private insurance plans is imminent. Individuals and households will soon be able to purchase coverage in insurance exchanges. These exchanges, depending on how they are designed, are expected to better pool risks and contain costs. Consequently, they may reduce the cost of premiums and out-of-pocket expenses for the average participant, like farm households, who have heretofore been relying on directly purchased private insurance plans.

Acknowledgments

A.K.M. is a professor, Department of Agricultural Economics and Agribusiness, Louisiana State University AgCenter and Louisiana State University, Baton Rouge, LA 70803, USA. H.S.E.-O. and M.C.A. are economist and senior economist in Farm Household Well-Being Branch, Resource & Rural Economics Division of the Economic Research Service, respectively. The views expressed here are not necessarily those of the Economic Research Service or the U.S. Department of Agriculture. We thank three anonymous reviewers for insightful comment and suggestions.

References

The Access Project, 2008. 2007 Health Insurance Survey of Farm and Ranch Operators. Issue Brief. No. 3, September, Boston, MA.
 Acquavella, J., Olsen, G., 1998. Cancer among farmers: a meta-analysis. *Ann. Epidemiol.* 8, 64–74.
 Arrow, K.J., 1963. Uncertainty and the welfare economics of medical care. *Am. Econ. Rev.* 53(5), 941–973.

Blair, A., Dosemeci, M., Heineman, E.F., 1993. Cancer and other causes of death among male and female farmers from twenty-three states. *Am. J. Ind. Med.* 23, 729–742.
 Blough, D.K., Madden, C., Hornbrook, M., 1999. Modelling risk using generalized linear models. *J. Health Econ.* 18(7), 153–171.
 Bowman, A., Azzalini, A., 1997. *Applied smoothing techniques for data analysis*. Clarendon Press, Oxford.
 Breusch, T.S., Pagan, A.R., 1979. A simple test for heteroscedasticity and random coefficient variation. *Econometrica* 47, 1287–1294.
 Buntin, M.B., Zaslavsky, A.M., 2004. Too much Ado about two-part models and transformation? Comparing methods of modeling Medicare expenditures. *J. Health Econ.* 23, 525–542.
 Cameron, A.C., Trivedi, P.K., Milne, F., Piggott, J., 1988. A microeconomic model of the demand for health care and health insurance in Australia. *Rev. Econ. Stud.* 55(1), 85–106.
 Cook, R.D., Weisberg, S., 1983. Diagnostics for heteroscedasticity in regression. *Biometrika* 70, 1–10.
 Cox, N.J. 2004. Speaking stata: graphing distributions. *Stata J.* 4(1), 66–88.
 Currie, J., Madrian, B., 1999. Health, health insurance and the labor market. In: Ashenfelter, O., Card, D. (Eds.), *Handbook of Labor Economics*, 3C. Elsevier, Amsterdam, pp. 3309–3406.
 Decker, C.S., 2009. Voluntary pollution control and local economic conditions as determinants of environmental monitoring and enforcement: evidence from the 1990s. *J. Appl. Econ. Policy* 28(Spring), 34–51. Available at: <http://www.wku.edu/jaep/html/documents/JAEPVol28%281%292009.pdf>
 Doty, M., Collins, S.R., Rustgi, S.D., Kriss, J.L., 2008. Seeing Red: The Growing Burden of Medical Bills and Debt Faced by U.S. Families, 42. The Commonwealth Fund Issue Brief, pub. 1164.
 Dow, W.H., Norton E.C., 2003. Choosing between and interpreting the Heckit and two-part models for corner solutions. *Health Serv. Outcomes Res. Methodol.* 4, 5–18.
 Duan, N., 1983. Smearing estimate: a nonparametric retransformation method. *J. Am. Stat. Assoc.* 78(386), 605–610.
 Duan, N., Manning, W.G., Morris, C.N., Newhouse, J.P., 1984. Choosing between the sample-selection model and the multipart model. *J. Bus. Econ. Stat.* 2, 283–289.
 Duan, N., Manning, W.G. Jr., Morris, C.N., Newhouse, J.P., 1983. A comparison of alternative models for the demand for medical care. *J. Bus. Econ. Stat.* 1, 115–126.
 Dubman, R.W., 2000. Variance estimation with USDA's farm costs and returns surveys and agricultural resource management study surveys. ERS Staff Paper, AGES 00-01.
 Feldstein, M.S., 1973. The welfare loss of excess health insurance. *J. Pol. Econ.*, 81(2), 251–80.
 Frondel, M., Vance, C., 2010. Driving for fun? Comparing the effects of fuel prices on weekend fuel consumption. *Energy Econ.* 32, 102–109.
 Gilleskie, D.B., Mroz, T.A., 2004. A flexible approach for estimating the effect of covariates on health expenditure. *J. Health Econ.* 23, 391–418.
 Greene, W.H., 2008. *Econometric Analysis*, 6th edition. Prentice-Hall, Englewood Cliffs, NJ.
 Gruber, J., Madrian, B., 2002. Health insurance, labor supply and job mobility: a critical review of literature. NBER working paper 8817. NBER, Cambridge, MA.
 Gruber, J., Poterba, J., 1994. Tax incentives and the decision to purchase health insurance: evidence from self-employed. *Q. J. Econ.* 109(3), 701–733.
 Hay, J.W., Olsen, R.J., 1984. Let them eat cake: a note on comparing alternative models of the demand for medical care. *J. Bus. Econ. Stat.* 2, 279–282.
 Hornbrook, M.C., Goodman, M.J., 1995. Assessing relative health plan risk with the RAND-36 health survey. *Inquiry* 32, 56–74.
 Hopkins, S., Kidd, M., 1996. The determinants of the demand for private health insurance under Medicare. *Appl. Econ.* 28, 1623–1632.
 Huber, P.J., 1967. The behavior of maximum likelihood estimates under non-standard conditions. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1 University of California Press, Berkeley, CA, pp. 221–23.

- Hurd, M.D., McGarry, K., 1997. Medical insurance and the use of health care services by elderly. *J. Health Econ.* 16(2), 129–154.
- Kott, P.S., 1997. Using the delete-a-group jackknife variance estimator in NASS surveys. U.S. Dept. Agr., National Agricultural Statistics Service.
- Manning, W.G., 1998. The logged dependent variable, heteroscedasticity, and the retransformation problem. *J. Health Econ.* 17, 283–295.
- Manning, W.G., Basu, A., Mullahy, J., 2005. Generalized modeling approach to risk adjustment of skewed outcomes data. *J. Health Econ.* 24, 465–488.
- Manning, W.G., Duan, N., Rogers, W.H., 1987. Monte Carlo evidence on the choice between sample selection and two-part models. *J. Econometrics* 35, 58–87.
- Manning, W.G., Morris, C.N., Newhouse, J.P., et al., 1981. A two-part model of the demand for medical care: preliminary results from the Health Insurance Study. In: van der Gaag, J., Perlman, M. (Eds.), *Health, Economics, and Health Economics*. North Holland, Amsterdam, pp. 103–123.
- Manning, W.G., Mullahy, J., 2001. Estimating log models: to transform or not to transform? *J. Health Econ.* 20, 461–494.
- McDonnell, K., Fronstin, P., 1999. *EBRI Health Databook*. Employee Benefit Research Institute, Washington, DC.
- McNamara, P.E., 2001. Health Insurance Coverage of Illinoisans: An Analysis of the Current Situation, Trends, and Correlated Health Behavior Using BRFSS Data. Illinois Department of Public Health-Illinois Assembly on the Uninsured, 4 October, pp. 1–31.
- McNamara, P.E., Straub, L., 2002. Performance Assessment of the Illinois Critical Access Hospital Program. Illinois Department of Public Health, Center for Rural Health, 1–23 December.
- Meyerhoefer, C., Pylypchuk, Y., 2008. Does participation in the food stamp program increase the prevalence of obesity and health care spending? *Am. J. Agric. Econ.* 90(2), 287–305.
- Mishra, A., El-Osta, H., Morehart, M., Johnson, J., Hopkins, J., 2002. Income, wealth, and the economic well-being of farm households. *AER* 812, ERS, USDA.
- Mullahy, J., 1998. Much Ado about two: Reconsidering retransformation and the two part model in health econometrics. *J. Health Econ.* 17, 247–281.
- Ngui, M., Burrows, C., Brown, K., 1990. Health insurance choice: an econometric analysis of ABS health and health demand insurance surveys. In Selby, S.C. (Ed.), *Economics and Health: 1989 Proceedings of the Eleventh Australian Conference of Health Economists*, Public Sector Management Institute, Monash University, Melbourne.
- Norton, E.C., Dow, W.H., Do, Y.K., 2008. Specification tests for the sample selection and two-part models. *Health Serv. Outcomes Res. Methods* 8, 201–208.
- Phelps, C.E., 1973. Demand for health insurance: a theoretical and empirical investigation. Rand Corporation, R-1054-OEO.
- Park, R., 1966. Estimation with heteroscedastic error terms. *Econometrica* 34(3), 888.
- Propper, C., 1989. An econometric analysis of the demand for private health insurance in England and Wales. *Appl. Econ.* 21, 777–792.
- Ramsey, J.B., 1969. Tests for specification errors in classical linear least squares regression analysis. *J. R. Statist. Soc. B* 31(2), 350–371.
- Savage, E., Wright, D., 1999. *Health Insurance and Health Care Utilization: Theory and Evidence from Australia 1989–90*. Mimeograph B2-Mimeograph. University of Sydney, Sydney.
- Shen, H., Brown, L.D., Zhi, H., 2006. Efficient estimation of log-normal means with application to pharmacokinetic data. *Stat. Med.* 25, 3023–3038.
- U.S. Bureau of the Census, 2008. Income, poverty, and health insurance coverage of people in the United States: 2007. Current Population Reports, Series P60-233, <http://www.census.gov/prod/2008pubs/p60-235.pdf>.
- U.S. Congress, Joint Committee on Taxation, 2008. Estimates of federal tax expenditures for fiscal years 2007–2011. 110th Congress, 1st Session. GPO, Washington, D.C., 24 September.
- U.S. Department of Health and Human Services, 2009. Family-level expenditures on health care and insurance premiums among the U.S. nonelderly population, 2006. Agency for Healthcare Research and Quality, March, MEPS Research Findings #29.
- U.S. Department of Agriculture. Economic Research Service, 2008a. Agriculture and Resource Management Survey. Briefing Room, <http://www.ers.usda.gov/Briefing/ARMS/resourceregions/resourceregions.htm>
- U.S. Department of Agriculture. Economic Research Service, 2008b. Farm household economics and well-being. Briefing Room, <http://www.ers.usda.gov/Briefing/WellBeing/>
- Van De Ven, W.P.M.M., Van Praag, B.M.S., 1981. Demand for deductibles in private health insurance. *J. Econometrics* 17(2), 229–252.
- White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48, 817–830.
- Wooldridge, J.M., 2002. *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, MA.
- Zheng, X., Zimmer, D.M., 2008. Farmers' health insurance and access to health care. *Am. J. Agric. Econ.* 90(1), 267–279.